

## Implementando un DataWarehouse.

Carmen Wolff

### Comenzando A Construir Un DW.

Para llevar a cabo con éxito un proyecto Datawarehouse, es vital considerar al inicio de su construcción tres factores esenciales: RRHH, Tecnología y Disciplina.

La disciplina es fundamental para el desarrollo del DW. Estas disciplinas son usadas para asegurar calidad, lograr sinergia, y mejorar el equipo de trabajo durante todo el proceso de desarrollo. Así los siguientes factores resultan ser imprescindibles para llevar a cabo la implementación de un DW:

- Prácticas de trabajo efectivo en el equipo de trabajo participante en el proyecto para lograr metas compartidas.
- Estándares, convenciones de calidad y en general la forma de resultados.
- Una metodología de desarrollo la cual defina formalmente los pasos y resultados del desarrollo.

### Análisis del Problema

La complejidad en el desarrollo se ha presentado como la principal desventaja de un DW. Esto se debe a que la realidad para cada negocio es distinta, y un DW debe responder a las características particulares que presenta cada uno de ellos, tanto de configuración como del conjunto de requisitos a satisfacer; por lo cual no es fácil estandarizar la forma de desarrollar este tipo de proyectos.

El empleo de una forma de trabajo ordenada es un factor de importancia en el desarrollo e implantación de proyectos de Datawarehousing, y la tendencia en general busca lograr a través del uso de una metodología, recortar los tiempos de desarrollo y programar la inversión de recursos de manera eficiente; además proporciona un lenguaje común logrando que exista comunicación, permitiendo la incorporación de nuevos miembros al equipo de trabajo siendo productivos inmediatamente.

En la actualidad no podemos asegurar cuál estrategia de implementación es mejor o peor, sin embargo al analizar las tendencias generales del mercado se encuentra que la estrategia de desarrollo de datamarts está siendo adoptada con mayor frecuencia en los últimos tiempos. A esta tendencia general se le ha identificado como la aproximación que garantiza la probabilidad de éxito más grande en la implantación de datawarehousing, tanto por la rapidez en la obtención de resultados en períodos cortos con inversiones moderadas como por la modularidad posible de alcanzar con este enfoque considerando cada datamart como un incremento del sistema final (datawarehouse).

### Construcción de un DW.

Al iniciar un proyecto Datawarehousing no debemos olvidar establecer un marco de referencia de construcción del DW.

Podemos distinguir en dicha construcción dos etapas principales: la definición de una arquitectura DW y la construcción de los incrementos DW.

#### **Definición de una arquitectura.**

*Arquitectura DW:*

Arquitectura DW establece el marco de trabajo, estándares y procedimientos para el DW a un nivel

empresarial. Los objetivos de las actividades de la arquitectura son simples, integrar al DW las necesidades de información empresarial.

#### *Resultados de la Arquitectura*

Los principales resultados del desarrollo de la arquitectura DW incluyen:

- El modelo de datos fuente.
- El modelo de datos conceptual DW.
- Arquitectura tecnológica DW.
- Estándares y procedimientos DW.
- El plan de implementación incremental para el DW.

Los modelos de datos proveen una estructura para identificar, nombrar, describir y asociar los componentes de una base de datos. En general se necesitan modelos de datos para los datos fuente como para los datos seleccionados para existir en el DW.

Los estándares DW son una parte importante de la arquitectura DW. Sin estándares, oportunidades para reusar no son posibles y hay riesgos de que partes del desarrollo no ganen trabajando juntos.

El plan de implementación DW es la parte de la arquitectura de DW que identifica los incrementos del DW y describe la secuencia de desarrollo de estos incrementos.

#### **Construcción en Incrementos, Datamarts.**

##### *Construcción Incremental:*

La implementación en incrementos warehouse desarrolla y genera un subconjunto del DW total. La implementación incremental es una aproximación pragmática para construir un warehouse a un nivel empresarial en forma evolutiva.

##### *Resultados de Implementación:*

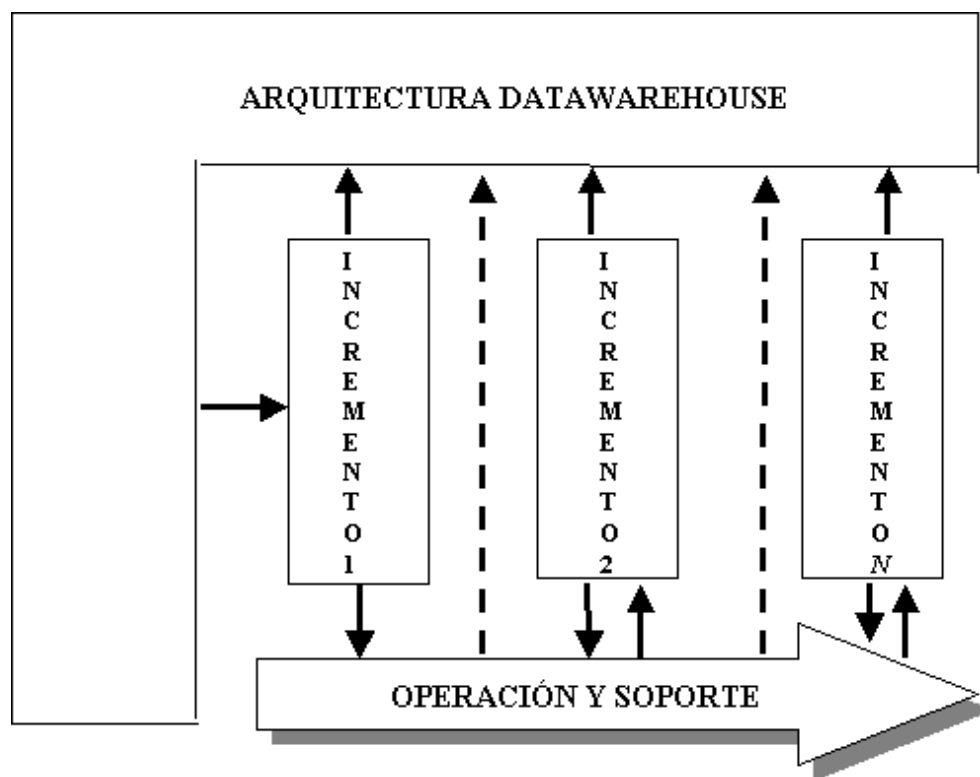
Los resultados más significativos de la implementación de un incremento DW, incluyen:

- Las bases de datos warehouse.
- Programas y procedimientos para extraer, transformar y cargar datos.
- Instalar herramientas de acceso a los datos.
- Poblar el DW con los datos necesarios.
- Poblar el catálogo de metadatos con los datos necesarios.
- Técnicas de uso y soporte el almacén

#### **Consideraciones de Implementación mediante DataMarts.**

- La arquitectura Datawarehouse se debe desarrollar al principio del proyecto.
- El primer incremento se desarrolla basándose en la arquitectura.
- La construcción del primer incremento puede causar cambios en la arquitectura.
- La operación del DW puede implicar la realización de cambios en la arquitectura.
- Cada incremento adicional puede extender el datawarehouse.
- Cada nuevo incremento puede causar ajustes en la arquitectura.
- La operación continua puede causar ajustes en la arquitectura.

La siguiente figura, intenta esquematizar el proceso de construcción del Datawarehouse.



*Figura 1 : Esquema de Construcción del DW[MicroSt96]*

## Metodología De Desarrollo de DW

De acuerdo a la forma de implementación analizada, se deben considerar y asociar metodologías congruentes con el desarrollo de incrementos dirigidos a grupos específicos en las organizaciones.

Al acercarse a la implementación de un DW con un conjunto de proyectos pequeños, altamente enfocados para implementar partes DW, resulta natural pensar en una metodología incremental para abordar su desarrollo. Pero no debemos olvidar la integración de cada incremento a la arquitectura Datawarehouse, así entonces el desarrollo evolutivo resulta ser una aproximación práctica de construcción de un DW.

De esta manera, estos proyectos pueden aprovechar los beneficios de la implementación incremental, que incluyen la contención de riesgos, oportunidades para aprender a desarrollar datamarts, entrega frecuente y la minimización del impacto sobre la comunidad de usuarios, y a la vez pueden ser organizados en forma secuencial, paralela o en una combinación de estructuras en serie y paralelo.

En tanto, el desarrollo evolutivo implica que cada vez que un incremento sea entregado, se debe operar y desarrollar simultáneamente el DW. De esta forma se logra integrar cada incremento a la estructura final DW.

El costo del desarrollo evolutivo queda dado por el incremento en la frecuencia de los cambios y la necesidad intensificada de realizar soporte del DW.

## Evolución del DW

### Evolución de la Implementación

La aproximación del desarrollo incremental es por naturaleza evolutiva. El primer incremento libera un subconjunto del DW el cual satisface un conjunto limitado de necesidades de información. Con cada incremento que es agregado el DW se vuelve más completo, quedando habilitado para satisfacer un mayor conjunto de necesidades de información.

### **Evolución de la Arquitectura**

El desarrollo incremental también ofrece una oportunidad para aprender y minimizar el impacto de cometer errores en el proceso de construcción; es poco probable que algún desarrollo de arquitectura de DW sea perfecto antes de construir su primer incremento. Ambas, las actividades de construcción de incrementos y las actividades de operación del DW, proveen retroalimentación valiosa que ayuda a refinar la arquitectura.

### **Conclusiones y comentarios a la serie de artículos**

En general, el factor que más incide en la necesidad de implementar un DW es la excesiva complejidad de los modelos relacionales que soportan los sistemas operacionales, lo cual dificulta en extremo la extracción de información para llevar a cabo labores de gestión de negocios. Y también es precisamente esta complejidad en los sistemas operacionales el factor que hace más complejo y tedioso el desarrollo de un Datawarehouse.

En el caso de tener que decidir las herramientas de desarrollo, esto pasa no sólo por análisis de precio, soporte y las clásicas consideraciones de hardware y software disponibles. También son factores claves a considerar el objetivo final (¿se desea construir DataMart independientes o a futuro se espera llegar a una integración de ellos?) como la facilidad de uso que otorgue al usuario.

Con las características que poseen las herramientas de desarrollo la labor de construcción en el ámbito multidimensional se agiliza enormemente. El esfuerzo principal entonces está en el diseño del modelo a implementar. Un buen diseño no sólo optimiza el tiempo de respuesta y el espacio físico utilizado, sino que permite además un crecimiento en el análisis (es decir, consultas no consideradas al inicio) a través de fórmulas y cálculos definidos en tiempos de ejecución.

Es imposible evaluar con eficacia cualquier diseño de datamart hasta que se carguen los datos y se muestren a los usuarios. Diseñar las bases de datos para los datamarts es totalmente diferente del diseño para los sistemas operacionales. El diseño de base de datos operacionales captura los requisitos del negocio y el proceso de diseño se considera un éxito si representa exactamente estos requisitos. En cambio para los datamarts, el proceso de diseño puede ser considerado acertado sólo si el diseño de base de datos refleja exactamente al usuario junto a sus requisitos procedurales, acomodando para ello datos de unos o más sistemas operacionales. La única forma de alcanzar una confiabilidad en el diseño del datamart es crear un ciclo de retroalimentación temprano en el proyecto. Este ciclo de retroalimentación utiliza a usuarios finales para evaluar los prototipos construidos, en un proceso iterativo que continúa a lo largo del proyecto. De esta manera el equipo de proyecto puede vigilar la aceptación de usuario y descubrir defectos de diseño en forma temprana.

### **Conclusiones y Comentarios Finales.**

La última meta de un proyecto DW es que el personal de la empresa lo utilice para satisfacer sus propias necesidades de información. Así las herramientas de acceso de datos son un componente esencial del DataWarehouse.

El diseño de un DW deberá necesariamente estar definido en forma menos precisa que el diseño de sistemas operacionales. Esto se debe a que estos últimos automatizan procesos de negocios cuyas reglas de negocio son más estables a lo largo del tiempo, lo que conlleva a una mejor definición de sus requerimientos; un DW en cambio, está orientado a mejorar el proceso de toma de decisiones, el cual

resulta ser un proceso muy variable a través del tiempo debido a las diversas y cambiantes situaciones en las cuales se deben analizar los datos, haciendo de su diseño un proceso definido en una forma mucho menos precisa.

Es muy importante considerar que los profesionales informáticos que participen en el proyecto, tengan un conocimiento del tema de negocios que contemplará existiendo un trabajo en conjunto con los usuarios finales de la aplicación, esto debido a que se debe tener por lo menos una proyección de los requerimientos futuros para poder darle un cierto nivel de flexibilidad a la estructura dimensional.

La experiencia tempranamente muestra que el desarrollo de un proyecto DW queda mejor dado en incrementos en una serie de pequeños proyectos, con cada uno de ellos agregando un tema a la estructura DW como parte de la arquitectura planeada.

Los sistemas DSS /EIS son limitados en el alcance de información que ellos proveen. El valor completo de un DW está dado cuando los sistemas DSS/EIS son usados en combinación con destrezas de acceso directamente al DW para encontrar sus necesidades especiales de información.

En este tipo de trabajos el ingeniero en informática actúa como un intermediario entre la labor de gestión propiamente y el desarrollo que podría realizar un analista de informática. Es un hecho que los ejecutivos en general, no pueden tomar decisiones sin la información adecuada, y a su vez el analista de informática no puede desarrollar una aplicación de apoyo a la gestión sin tener los conocimientos de toma de decisiones adecuados. El analista debe saber cómo van a ser usados los datos, qué tipo de consultas se realizarán, qué significa la relación entre un dato y otro, etc. Es por lo tanto el objetivo del ingeniero el realizar el modelo y la estructura que sustente la base que permita al analista construirla, y a los ejecutivos consultarla.

Finalmente, no se puede dejar de mencionar el gran aporte que una tecnología como esta significa para los usuarios que son, al fin y al cabo quienes tienen bajo su responsabilidad la labor de gestión de la empresa, decidiendo no sólo el futuro de su empresa sino también del propio desarrollo del país y porque no decirlo, de la humanidad.

Es posible prever la implementación de esta tecnología a nivel masivo por parte de las grandes empresas. El hecho de hacer hincapié respecto a las grandes empresas se debe básicamente a dos factores: debido al alto costo de inversión que tiene esta tecnología y por otra parte, debido al volumen de datos alcanzado para estas empresas que resultan necesarios para desempeñar las labores de gestión.

## Bibliografía

- [Communi98] Communications of the ACM, September 1998, Volume41 - Number9 : Datawarehous
- [Compute96A] Computers World N°115 , Septiembre 1996.
- [Compute96B] Computer World N°119, Noviembre 1996.
- [Compute97A] Computer World 11 Junio 1997.
- [Compute97B] Computer World N°135, 25 junio 1997.
- [Compute97C] Computer World N°136, 7 Julio 1997.
- [Compute97D] Computer World N°139, 20 Agosto 1997.
- [Compute97E] Computer World N°145, 12 Noviembre 1997.

- [Compute98] Computer World N°151 18 Febrero 1998.
- [Corey93] "Oracle Data Warehousing", Michael J. Corey & Michael Abbey, Computer World 1993 pág. 218.
- [Informa96] Revista Informática Volumen 18, 8 Septiembre 1996.
- [Mcguff1] "Designing The Perfect Datawarehouse", Frank McGuff, <http://www.techguide.com/>
- [Mcguff2] "Datawarehouse Modeling", Frank McGuff, <http://www.techguide.com/>
- [MicroSt96] "Data Warehousing, Data Modeling and Design", MicroStrategy Education, Nov 96.
- [Paper1] "Managing The Warehouse Throughout Its Lifecycle", <http://www.techguide.com/>
- [Paper2] "Building A Decision Support Architecture For Datawarehousing", <http://www.techguide.com/>
- [Paper3] "Putting Metadata To Work In The Warehouse" , <http://www.techguide.com/>
- [Paper4] "Enterprise Storage: Delivering Data Warehousing Business Results", <http://www.techguide.com/>
- [Paper5] "A Practical Guide To Getting Started With Data Warehousing", <http://www.techguide.com/>
- [Solucio96] Especial de Datawarehouse, Revista Soluciones Avanzadas, Edición Junio 1996 - Pág. 72.
- [Softwar96] "Executive Briefing for the Datawarehouse", Software AG Education Services, January 1996
- [Wolff98A] "Datawarehousing", Taller De Computación - Tópico: Bases De Datos, Depto. Ingenieria Informática Y Ciencias De La Computación - Facultad De Ingeniería - Universidad De Concepción, Desarrollado por Carmen Gloria Wolff
- [Wolff98B] "Consideraciones para Enfrentar el Desarrollo de un DW", Asignatura Gestión Project Ingeniería de Software (GPIS), Depto. Ingeniería Informática Y Ciencias De La Computación - Facultad De Ingeniería -, Preparado por Carmen Gloria Wolff.

## Direcciones Internet Recomendadas

<http://www.techguide.com/>

<http://pwp.starnetinc.com/savmony.html>

<http://www.people.memphis.edu/~tsakagch/dw-web.htm>

<http://www.datamation.com>

<http://www.dmreview.com>

<http://www.dbdp.com>

<http://www.dbmsmag.com>

## Glosario

**Agregación** : Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del DataWarehouse en forma más rápida y fácil.

**Datawarehouse** : Base de datos que almacena una gran cantidad de datos transaccionales integrados para ser usados para análisis gestionales por usuarios especializados (tomadores de decisión de la empresa).

**DataMart** : Conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro DataMart.

**Dataminig** : Análisis de los datos para descubrir relaciones, patrones, o asociaciones desconocidas.

**Diccionario de Datos** : Un compendio de definiciones y especificaciones para las categorías de datos y sus relaciones.

**Dimensión** : Entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.

**Drill Down** : Exponer progresivamente más detalle (dentro de un reporte o consulta), mediante selecciones de ítemes sucesivamente.

**Drill-Up** : Es el efecto contrario a drill-down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.

**DSS** : Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

**ETT (Extracción, Transformación y Transporte de datos)** : Pasos por los que atraviesan los datos para ir desde el sistema OLTP (o la fuente de datos utilizada) a la bodega dimensional. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos. Finalmente está el Transporte, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un Data Mart no es necesario atravesar por todos estos pasos, pues al ser información localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

**Jerarquía** : Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía. Por ejemplo : los Centros de Responsabilidad están asociados a un Tipo de Unidad, el cual pueden corresponder a una gerencia, subgerencia, superintendencia, etc.; por otro parte, cada CR está asociado a otro CR a nivel administrativo y, también existe una clasificación a nivel funcional.

**Limit** : Comando propio del lenguaje Express, que permite seleccionar los datos a visualizar. Limita el acceso a los datos dejando 'invisible' o no accesible el resto de ellos.

**Olap (On-line Analytical Processing)** : Conjunto de principios que proveen una ambiente de trabajo dimensional para soporte decisional.

**Oltp (On-line Transaction Processing)** : Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.

**Rollup** : Comando propio del lenguaje Oracle Express, que simboliza las sumas agregadas de una variable a través de los niveles jerárquicos de las dimensiones que la sustentan.

**Snapshot** : Imagen instantánea de los datos en un tiempo dado.

**Sumarización** : Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

**Tabla Dimensional** : Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.